



#2

OIPE

## RAW SEQUENCE LISTING

DATE: 01/26/2002

PATENT APPLICATION: US/09/996,223

TIME: 14:13:14

Input Set : N:\Crf3\RULE60\09996223.raw

Output Set: N:\CRF3\01252002\I996223.raw

## SEQUENCE LISTING

## 3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: Hillman, Jennifer L.

6 Lal, Preeti

7 Corley, Neil C.

8 Shah, Purvi

11 (ii) TITLE OF INVENTION: HUMAN CITRATE SYNTHASE HOMOLOG

13 (iii) NUMBER OF SEQUENCES: 3

15 (iv) CORRESPONDENCE ADDRESS:

16 (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.

17 (B) STREET: 3174 Porter Drive

18 (C) CITY: Palo Alto

19 (D) STATE: CA

20 (E) COUNTRY: USA

21 (F) ZIP: 94304

23 (v) COMPUTER READABLE FORM:

24 (A) MEDIUM TYPE: Diskette

25 (B) COMPUTER: IBM Compatible

26 (C) OPERATING SYSTEM: DOS

27 (D) SOFTWARE: FastSEQ for Windows Version 2.0

29 (vi) CURRENT APPLICATION DATA:

C--&gt; 30 (A) APPLICATION NUMBER: US/09/996,223

C--&gt; 31 (B) FILING DATE: 27-Nov-2001

32 (C) CLASSIFICATION:

34 (vii) PRIOR APPLICATION DATA:

35 (A) APPLICATION NUMBER: 08/970,134

36 (B) FILING DATE:

38 (viii) ATTORNEY/AGENT INFORMATION:

39 (A) NAME: Billings, Lucy J.

40 (B) REGISTRATION NUMBER: 36,749

41 (C) REFERENCE/DOCKET NUMBER: PF-0425 US

43 (ix) TELECOMMUNICATION INFORMATION:

44 (A) TELEPHONE: 650-855-0555

45 (B) TELEFAX: 650-845-4166

46 (C) TELEX:

49 (2) INFORMATION FOR SEQ ID NO: 1:

51 (i) SEQUENCE CHARACTERISTICS:

52 (A) LENGTH: 466 amino acids

53 (B) TYPE: amino acid

54 (C) STRANDEDNESS: single

55 (D) TOPOLOGY: linear

57 (vii) IMMEDIATE SOURCE:

58 (A) LIBRARY: ENDCNOT03

ENTERED

## RAW SEQUENCE LISTING

DATE: 01/26/2002

PATENT APPLICATION: US/09/996,223

TIME: 14:13:14

Input Set : N:\Crf3\RULE60\09996223.raw

Output Set: N:\CRF3\01252002\I996223.raw

```

59          (B) CLONE: 2171653
61      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
63 Met Ala Leu Leu Thr Ala Ala Ala Arg Leu Leu Gly Thr Lys Asn Ala
64   1          5          10          15
65 Ser Cys Leu Val Leu Ala Ala Arg His Ala Ser Ala Ser Ser Thr Asn
66          20          25          30
67 Leu Lys Asp Ile Leu Ala Asp Leu Ile Pro Lys Glu Gln Ala Arg Ile
68          35          40          45
69 Lys Thr Phe Arg Gln Gln His Gly Lys Thr Val Val Gly Gln Ile Thr
70          50          55          60
71 Val Asp Met Met Tyr Gly Gly Met Arg Gly Met Lys Gly Leu Val Tyr
72          65          70          75          80
73 Glu Thr Ser Val Leu Asp Pro Asp Glu Gly Ile Arg Phe Arg Gly Phe
74          85          90          95
75 Ser Ile Pro Glu Cys Gln Lys Leu Leu Pro Lys Ala Lys Gly Gly Glu
76          100          105          110
77 Glu Pro Leu Pro Glu Gly Leu Phe Trp Leu Leu Val Thr Gly His Ile
78          115          120          125
79 Pro Thr Glu Glu Gln Val Ser Trp Leu Ser Lys Glu Trp Ala Lys Arg
80          130          135          140
81 Ala Ala Leu Pro Ser His Val Val Thr Met Leu Asp Asn Phe Pro Thr
82          145          150          155          160
83 Asn Leu His Pro Met Ser Gln Leu Ser Ala Ala Val Thr Ala Leu Asn
84          165          170          175
85 Ser Glu Ser Asn Phe Ala Arg Ala Tyr Ala Gln Gly Ile Ser Arg Thr
86          180          185          190
87 Lys Tyr Trp Glu Leu Ile Tyr Glu Asp Ser Met Asp Leu Ile Ala Lys
88          195          200          205
89 Leu Pro Cys Val Ala Ala Lys Ile Tyr Arg Asn Leu Tyr Arg Glu Gly
90          210          215          220
91 Ser Gly Ile Gly Ala Ile Asp Ser Asn Leu Asp Trp Ser His Asn Phe
92          225          230          235          240
93 Thr Asn Met Leu Gly Tyr Thr Asp His Gln Phe Thr Glu Leu Thr Arg
94          245          250          255
95 Leu Tyr Leu Thr Ile His Ser Asp His Glu Gly Gly Asn Val Ser Ala
96          260          265          270
97 His Thr Ser His Leu Val Gly Ser Ala Leu Ser Asp Pro Tyr Leu Ser
98          275          280          285
99 Phe Ala Ala Ala Met Asn Gly Leu Ala Gly Pro Leu His Gly Leu Ala
100          290          295          300
101 Asn Gln Glu Val Leu Val Trp Leu Thr Gln Leu Gln Lys Glu Val Gly
102          305          310          315          320
103 Lys Asp Val Ser Asp Glu Lys Leu Arg Asp Tyr Ile Trp Asn Thr Leu
104          325          330          335
105 Asn Ser Gly Arg Val Val Pro Gly Tyr Gly His Ala Val Leu Arg Lys
106          340          345          350
107 Thr Asp Pro Arg Tyr Thr Cys Gln Arg Glu Phe Ala Leu Lys His Leu
108          355          360          365
109 Pro Asn Asp Pro Met Phe Lys Leu Val Ala Gln Leu Tyr Lys Ile Val

```

## RAW SEQUENCE LISTING

DATE: 01/26/2002

PATENT APPLICATION: US/09/996,223

TIME: 14:13:14

Input Set : N:\Crf3\RULE60\09996223.raw

Output Set: N:\CRF3\01252002\I996223.raw

```

110      370      375      380
111 Pro Asn Val Leu Leu Glu Gln Gly Lys Ala Lys Asn Pro Trp Pro Asn
112 385      390      395      400
113 Val Asp Ala His Ser Gly Val Leu Leu Gln Tyr Tyr Gly Met Thr Glu
114      405      410      415
115 Met Asn Tyr Tyr Thr Val Leu Phe Gly Val Ser Arg Ala Leu Gly Val
116      420      425      430
117 Leu Ala Gln Leu Ile Trp Ser Arg Ala Leu Gly Phe Pro Leu Glu Arg
118      435      440      445
119 Pro Lys Ser Met Ser Thr Glu Gly Leu Met Lys Phe Val Asp Ser Lys
120      450      455      460
121 Ser Gly
122 465
124 (2) INFORMATION FOR SEQ ID NO: 2:
126 (i) SEQUENCE CHARACTERISTICS:
127 (A) LENGTH: 2131 base pairs
128 (B) TYPE: nucleic acid
129 (C) STRANDEDNESS: single
130 (D) TOPOLOGY: linear
132 (vii) IMMEDIATE SOURCE:
133 (A) LIBRARY: ENDCNOT03
134 (B) CLONE: 2171653
136 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
138 CGCCGGTTTCG TCTACTCTTT CCTTCAGCCG CCTCCTTTCA ACCTTGTC AA CCCGTCGGCG 60
139 CGGCCTCTGG TGCAGCGGCG GCGGCTCCTG TTCCTGCCGC AGCTCTCTCC CTTTCTTACC 120
140 TCCCCACCAG ATCCCGGAGA TCGCCCGCCA TGGCTTTACT TACTGCGGCC GCCCGGCTCT 180
141 TGGGAACCAA GAATGCATCT TGTCTTGTTT TTGCAGCCCG GCATGCCAGT GCTTCCTCCA 240
142 CGAATTGAA AGACATATTG GCTGACCTGA TACCTAAGGA GCAGGCCAGA ATTAAGACTT 300
143 TCAGGCAGCA ACATGGCAAG ACGGTGGTGG GCCAAATCAC TGTGGACATG ATGTATGGTG 360
144 GCATGAGAGG CATGAAGGGA TTGGTCTATG AAACATCAGT TCTTGATCCT GATGAGGGCA 420
145 TCCGTTTCCG AGGCTTTAGT ATCCCTGAAT GCCAGAAACT GCTACCCAAG GCTAAGGGTG 480
146 GGAAGAACC CCTGCCTGAG GGCTTATTTT GGCTGCTGGT AACTGGACAT ATCCCAACAG 540
147 AGGAACAGGT ATCTTGGCTC TCAAAGAGT GGGCAAAGAG GGCAGCTCTG CCTTCCCATG 600
148 TGGTCACCAT GCTGGACAAC TTTCCACCA ATCTACACCC CATGTCTCAG CTCAGTGCAG 660
149 CTGTTACAGC CCTCAACAGT GAAAGTAACT TTGCCCAGC ATATGCACAG GGTATCAGCC 720
150 GAACCAAGTA CTGGGAGTTG ATTTATGAAG ACTCTATGGA TCTAATCGCA AAGCTACCTT 780
151 GTGTTGCAGC AAAGATCTAC CGAAATCTCT ACAGAGAAGG CAGCGGTATT GGGGCCATTG 840
152 ACTCTAACCT GGA CTGGTCT CACAATTTCA CCAACATGTT AGGCTATACT GATCATCAGT 900
153 TCACTGAGCT CACGCGCCTG TACCTCACCA TCCACAGTGA CCATGAGGGT GGCAATGTAA 960
154 GTGCCCATAC CAGCCATTTG GTGGGCAGTG CCCTTTCCGA CCCTTACCTG TCCTTTGCAG 1020
155 CAGCCATGAA CGGGCTGGCA GGGCCTCTCC ATGGACTGGC AAATCAGGAA GTGCTTGTCT 1080
156 GGCTAACACA GCTGCAGAAG GAAGTTGGCA AAGATGTGTC AGATGAGAAG TTACGAGACT 1140
157 ACATCTGGAA CACACTCAAC TCAGGACGGG TTGTTCAGG CTATGGCCAT GCAGTACTAA 1200
158 GGAAGACTGA TCCGCGATAT ACCTGTGAGC GAGAGTTTGC TCTGAAACAC CTGCCTAATG 1260
159 ACCCATGTT TAAGTTGGTT GCTCAGCTGT ACAAGATTGT GCCAATGTG CTCTTAGAGC 1320
160 AGGGTAAAGC CAAGATCCT TGGCCCAATG TAGATGCTCA CAGTGGGGTG CTGCTCCAGT 1380
161 ATTATGGCAT GACGGAGATG AATTACTACA CGGTCTGTG TGGGGTGTCA CGAGCATTGG 1440
162 GTGTACTGGC ACAGCTCATC TGGAGCCGAG CCTTAGGCTT CCCTCTAGAA AGGCCCAAGT 1500
163 CCATGAGCAC AGAGGGTCTG ATGAAGTTTG TGGACTCTAA GTCAGGGTAA AACTGGAGAC 1560

```

## RAW SEQUENCE LISTING

DATE: 01/26/2002

PATENT APPLICATION: US/09/996,223

TIME: 14:13:15

Input Set : N:\Crf3\RULE60\09996223.raw

Output Set: N:\CRF3\01252002\I996223.raw

```

164 TGGGTGAAAG TGACTACCAG AAAGTGAGGA AGCCTAAATA AAAAGTATAC TTTTGTTC 1620
165 GGGGGCCTTT AAAGACTTAA GATTAAATTA TATCTGAGGC ACTGATAATA TGTTTGAGGT 1680
166 TAAAAATATAA ATTAAGACTT TAAAAGATGA AAAATGGTCC CTTCTTCCCT AATCAGCTCC 1740
167 CTTCCCCTGC CTGGTATGAG TTGCCCATCA TACGCATGGT CCTGGAGGAT GACCAGGACT 1800
168 AATGCATGTG GTATGAGTAG GTTTGGCCCC CTCACTATCT CTAGAGTGAG AATCTGGCTC 1860
169 CTGTTTCCAT GGGTCAAAGC CGGTTGCAGA GAATCTGTAG TCACTTTGGA GCTTTAGCTT 1920
170 CTCTGCCAAG CCCTCAATAA GCCAGCAAAC CAGGACTCTG CCCCTTCTGT TTCCATAGGA 1980
171 ATCATGTTGG ATAGTCAGCT GTACCAAGCC CCTTGGCCCT CTCCCATGCA CACAAACACC 2040
172 TCCTAGCAAG ACCTGTTGGT TAGCTGGACA TGCTTTGGCA ATTTTTTTAT ACTACCAAGT 2100
173 GACCATATTG GCATGGCATT TTTTGGTGAT G 2131

```

175 (2) INFORMATION FOR SEQ ID NO: 3:

177 (i) SEQUENCE CHARACTERISTICS:

178 (A) LENGTH: 464 amino acids

179 (B) TYPE: amino acid

180 (C) STRANDEDNESS: single

181 (D) TOPOLOGY: linear

183 (vii) IMMEDIATE SOURCE:

184 (A) LIBRARY: GenBank

185 (B) CLONE: 164419

187 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

189 Met Ala Leu Leu Thr Ala Ala Ala Arg Leu Phe Gly Ala Lys Asn Ala
190 1 5 10 15
191 Ser Cys Leu Val Leu Ala Ala Arg His Ala Ser Ala Ser Ser Thr Asn
192 20 25 30
193 Leu Lys Asp Ile Leu Ala Asp Leu Ile Pro Lys Glu Gln Ala Arg Ile
194 35 40 45
195 Lys Thr Phe Arg Gln Gln His Gly Asn Thr Val Val Gly Gln Ile Thr
196 50 55 60
197 Val Asp Met Met Tyr Gly Gly Met Arg Gly Met Lys Gly Leu Val Tyr
198 65 70 75 80
199 Glu Thr Ser Val Leu Asp Pro Asp Glu Gly Ile Arg Phe Arg Gly Tyr
200 85 90 95
201 Ser Ile Pro Glu Cys Gln Lys Met Leu Pro Lys Ala Lys Gly Gly Glu
202 100 105 110
203 Glu Pro Leu Pro Glu Gly Leu Phe Trp Leu Leu Val Thr Gly Gln Ile
204 115 120 125
205 Pro Thr Glu Glu Gln Val Ser Trp Leu Ser Lys Glu Trp Ala Lys Arg
206 130 135 140
207 Ala Ala Leu Pro Ser His Val Val Thr Met Leu Asp Asn Phe Pro Thr
208 145 150 155 160
209 Asn Leu His Pro Met Ser Gln Leu Ser Ala Ala Ile Thr Ala Leu Asn
210 165 170 175
211 Ser Glu Ser Asn Phe Ala Arg Ala Tyr Ala Glu Gly Ile His Arg Thr
212 180 185 190
213 Lys Tyr Trp Glu Leu Ile Tyr Glu Asp Cys Met Asp Leu Ile Ala Lys
214 195 200 205
215 Leu Pro Cys Val Ala Ala Lys Ile Tyr Arg Asn Leu Tyr Arg Glu Gly
216 210 215 220
217 Ser Ser Ile Gly Ala Ile Asp Ser Lys Leu Asp Trp Ser His Asn Phe

```

## RAW SEQUENCE LISTING

DATE: 01/26/2002

PATENT APPLICATION: US/09/996,223

TIME: 14:13:15

Input Set : N:\Crf3\RULE60\09996223.raw

Output Set: N:\CRF3\01252002\I996223.raw

218	225		230		235		240
219	Thr Asn Met Leu Gly Tyr Thr Asp Ala Gln Phe Thr Glu Leu Met Arg						
220		245		250		255	
221	Leu Tyr Leu Thr Ile His Ser Asp His Glu Gly Gly Asn Val Ser Ala						
222		260		265		270	
223	His Thr Ser His Leu Val Gly Ser Ala Leu Ser Asp Pro Tyr Leu Ser						
224		275		280		285	
225	Phe Ala Ala Ala Met Asn Gly Leu Ala Gly Pro Leu His Gly Leu Ala						
226		290		295		300	
227	Asn Gln Glu Val Leu Val Trp Leu Thr Gln Leu Gln Lys Glu Val Gly						
228	305		310		315		320
229	Lys Asp Val Ser Asp Glu Lys Leu Arg Asp Tyr Ile Trp Asn Thr Leu						
230		325		330		335	
231	Asn Ser Gly Arg Val Val Pro Gly Tyr Gly His Ala Val Leu Arg Lys						
232		340		345		350	
233	Thr Asp Pro Arg Tyr Thr Cys Gln Arg Glu Phe Ala Leu Lys His Leu						
234		355		360		365	
235	Pro His Asp Pro Met Phe Lys Leu Val Ala Gln Leu Tyr Lys Ile Val						
236		370		375		380	
237	Pro Asn Val Leu Leu Glu Gln Gly Lys Ala Lys Asn Pro Trp Pro Asn						
238	385		390		395		400
239	Val Asp Ala His Ser Gly Val Leu Leu Gln Tyr Tyr Gly Met Thr Glu						
240		405		410		415	
241	Met Asn Tyr Tyr Thr Val Leu Phe Gly Val Ser Arg Ala Leu Gly Val						
242		420		425		430	
243	Leu Ala Gln Leu Ile Trp Ser Arg Ala Leu Gly Phe Pro Leu Glu Arg						
244		435		440		445	
245	Pro Lys Ser Met Ser Thr Asp Gly Leu Ile Lys Leu Val Asp Ser Lys						
246		450		455		460	

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/996,223

DATE: 01/26/2002

TIME: 14:13:16

Input Set : N:\Crf3\RULE60\09996223.raw

Output Set: N:\CRF3\01252002\I996223.raw

L:30 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:31 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]